Stat 100a: Introduction to Probability.

Outline for the day

1. Hand in HW3.
2. Review list.
3. Examples.
4. Tournaments.

Thu Dec 6 is the final exam, here in class, 11am to 12:15pm.

Again any notes and books are fine, and bring a pencil and a calculator.

Also bring your student ID to the exam.

1. HAND IN HW3. 2. Review list.

1) Basic principles of counting.
2) Axioms of probability, and addition rule.
3) Permutations & combinations.
4) Conditional probability.
5) Independence.
6) Multiplication rules. P(AB) = P(A) P(B|A) [= P(A)P(B) if ind.]
7) Odds ratios.
8) Random variables (RVs).
9) Discrete RVs, and probability mass function (pmf).
10) Expected value.
11) Pot odds calculations.
12) Luck and skill.
13) Variance and SD.
14) Bernoulli RV. [0-1. $\mu = p$, $\sigma = \sqrt{(pq)}$.]
15) Binomial RV. [# of successes, out of n tries. $\mu = np$, $\sigma = \sqrt{(npq)}$.]
16) Geometric RV. [# of tries til 1st success. $\mu = 1/p$, $\sigma = (\sqrt{q}) / p$.]
17) Negative binomial RV. [# of tries til rth success. $\mu = r/p$, $\sigma = (\sqrt{rq}) / p$.]
18) Poisson RV [# of successes in some time interval. [$\mu = \lambda$, $\sigma = \sqrt{\lambda}$.]
19) E(X+Y), V(X+Y) (ch. 7.1).
20) Bayes's rule (ch. 3.4).
20) Continuous RVs, Uniform, Normal, Exponential and Pareto.
21) Probability density function (pdf). Recall F'(c) = f(c), where F(c) = cdf.
22) Moment generating functions
23) Markov and Chebyshev inequalities
24) Law of Large Numbers (LLN) and Fundamental Theorem of Poker.
25) Central Limit Theorem (CLT)
26) Conditional expectation.
27) Confidence intervals for the sample mean and sample size calculations.
28) Random Walks, Reflection Principle, Ballot Theorem, avoiding zero
29) Chip proportions, doubling up, and induction.
30) Bivariate normal distribution and the conditional distribution of Y given X.
31) Covariance and correlation.

Basically, we've done all of ch. 1-7 except 6.7 and the optimal strategy stuff in 6.3.

P(You have AK | you have exactly one ace)?

= P(You have AK and exactly one ace) / P(exactly one ace)

= P(AK) / P(exactly one ace)

= (16/C(52,2)) ÷ (4x48/C(52,2))

= 4/48 = 8.33%.

P(You have AK | you have at least one ace)?

= P(You have AK and at least one ace) / P(at least one ace)

= P(AK) / P(at least one ace)

= (16/C(52,2)) ÷ (((4x48 + C(4,2))/C(52,2)) ~ 8.08%.

P(You have AK | your FIRST card is an ace)?

= 4/51 = 7.84%.

Suppose there are 1000 players in a casino. Each of them is playing holdem. What is the expected number of people who have pocket aces?

Let X_1 = 1 if player 1 has pocket aces, and 0 otherwise.

X_2 = 1 if player 2 has pocket aces, and 0 otherwise.

X_3 = 1 if player 3 has pocket aces, and 0 otherwise, etc.

X_1 and X_2 are not independent. Nevertheless, if Y = the number of people with AA,

then $Y = X_1 + X_2 + ... + X_{1000}$, and

$E(Y) = E(X_1) + E(X_2) + ... + E(X_{1000})$

$= C(4,2)/C(52,2) \times 1000$

$\sim 4.52.$

Let X = the number of queens you have and Y = the number of face cards you have. What is cov(X,Y)?

cov(X,Y) = E(XY) – E(X)E(Y).

$X = X_1 + X_2$, where $X_1 = 1$ if your first card is a queen and $X_2 = 1$ if your 2nd card is a queen,

so $E(X) = E(X_1) + E(X_2)$ = 1/13 + 1/13 = 2/13. Similarly, E(Y) = 3/13 + 3/13 = 6/13.

E(XY)? XY = 4 if you have QQ, or 2 if you have KQ or QJ, or 1 if you have Qx, where x is not a face card, or 0 otherwise. So E(XY) = 4 x C(4,2)/C(52,2) + 2 x (16+16)/C(52,2) + 1 x (4x40)/C(52,2) + 0 = 0.187.

So, cov(X,Y) = 0.187 – 2/13 x 6/13 =

= 0.116.

CLT Example

Suppose X1, X2, ..., X100 are 100 iid draws from a population with mean μ=70 and sd σ=10. What is the approximate distribution of the sample mean, $\bar{x}$?

By the CLT, the sample mean is approximately normal with mean μ and sd $\sigma/\sqrt{n}$, i.e. $\sim N(70, 1^2)$.

Now suppose Y1, Y2, ..., Y100 are iid draws, independent of X1, X2, X100, with mean μ=80 and sd σ=25. What is the approximate distribution of $\bar{x} - \bar{y} = Z$?

Now the sample mean of the first sample is approximately $N(70, 1^2)$ and similarly the negative sample mean of the 2nd sample is approximately $N(-80, 2.5^2)$, and the two are independent, so their sum Z is approximately normal.

Its mean is 70-80 = -10,

and $var(Z) = 1^2+2.5^2 = 7.25$, so $Z \sim N(-10, 2.69^2)$, because $2.69^2 = 7.25$.

Remember, if X and Y are ind., then var(X+Y) = var(X) + var(Y).

Suppose you have 10 players at the table. What is the expected number of players who have 2 face cards?

P(2 face cards) = C(12,2)/C(52,2) = 4.98%.

Let X1 = 1 if player 1 has 2 face cards, and X1 = 0 otherwise.

X2 = 1 if player 2 has 2 face cards, and X2 = 0 otherwise. etc.

X = ∑ Xi = total number of players with 2 face cards.

E(X) = ∑ E(Xi) = 10 x 4.98% = 0.498.

Let $X = N(0, 0.8^2)$ and $\varepsilon = N(0, 0.1^2)$ and independent of X. Let $Y = 5 + 0.3 X + \varepsilon$.

Find E(X), E(Y), E(Y|X), var(X), var(Y), cov(X,Y), and ρ = cor(X,Y).

$E(X) = 0$.

$E(Y) = E(5 + 0.3X + \varepsilon) = 5 + 0.3\, E(X) + E(\varepsilon) = 5$.

$E(Y|X) = E(5 + 0.3X + \varepsilon \mid X) = 5 + 0.3X + E(\varepsilon \mid X) = 5 + 0.3X$ since ε and X are ind.

$\mathrm{var}(X) = 0.64$.

$\mathrm{var}(Y) = \mathrm{var}(5 + 0.3\, X + \varepsilon) = \mathrm{var}(0.3X + \varepsilon) = 0.3^2\, \mathrm{var}(X) + \mathrm{var}(\varepsilon) + 2*0.3\, \mathrm{cov}(X,\varepsilon)$

$= 0.3^2(0.64) + 0.1^2 + 0 = 0.0676$.

$\mathrm{cov}(X,Y) = \mathrm{cov}(X, 5 + 0.3X + \varepsilon) = 0.3\, \mathrm{var}(X) + \mathrm{cov}(X, \varepsilon) = 0.3(0.64) + 0 = 0.192$.

$\rho = \mathrm{cov}(X,Y)/(\mathrm{sd}(X)\, \mathrm{sd}(Y)) = 0.192 / (0.8 \text{ x } \sqrt{.0676}) = 0.923$.

Suppose (X,Y) are bivariate normal with E(X) = 3, var(X) = 4, E(Y) = 5, var(Y) = 6, $\rho = 0.7$,

What is the distribution of Y given X = 8?

Given X = 8, Y is normal. Write $Y = \beta_1 + \beta_2 X + \varepsilon$ where ε is normal with mean 0, ind. of X.

Recall $\beta_2 = \rho\, \sigma_y/\sigma_x = 0.7 \times \sqrt{6}/2 = 0.857$.

So $Y = \beta_1 + 0.857\, X + \varepsilon$.

To get β_1, $5 = E(Y) = \beta_1 + 0.857\, E(X) + E(\varepsilon) = \beta_1 + 0.857\,(3) + 0$. So $5 = \beta_1 + 2.571$. $\beta_1 = 2.429$.

So $Y = 2.429 + 0.857\, X + \varepsilon$, where ε is normal with mean 0 and ind. of X.

What is var(ε)?

var(Y) = 6, and var(y) = var($2.429 + 0.857\, X + \varepsilon$) = 0.857^2 var(X) + var(ε) + 2(0.857) cov(X,ε)
= 0.857^2 (4) + var(ε) + 0. So var(ε) = $6 - 0.857^2(4) = 3.06$ and sd(ε) = $\sqrt{3.06} = 1.75$.

So $Y = 2.429 + 0.857\, X + \varepsilon$, where ε is $N(0, 1.75^2)$ and ind. of X.

Given X = 8, $Y = 2.429 + 0.857(8) + \varepsilon = 9.285 + \varepsilon$, so $Y|(X{=}8) \sim N(9.285, 1.75^2)$.

Suppose X and Y have joint density $f(x,y) = a(x+y^2)$, for x and y in (0,1), and $f(x,y) = 0$ otherwise.

What is a? What is the marginal density of X?

$1 = \int\int a(x+y^2)dydx = a \iint(x+y^2)dydx = a\int(xy + y^3/3)|_{y=0}^1\, dx = a\int(x + 1/3)dx = a\int x^2/2 + x/3\ |_{x=0}^1$

$= a(½ + 1/3) = a(5/6)$. So $a = 6/5$.

The marginal density $f(x) = \int f(x,y)\, dy = \int 6/5\ (x+y^2)dy = 6/5\ xy + 6/5\ y^3/3\ |_{y=0}^1 = 6/5\ x + 2/5$.

To verify this is a density, $\int f(x)\, dx = \int 6/5x + 2/5\ dx = 6/5\ x^2/2 + 2x/5\ |_{x=0}^1 = 3/5 + 2/5 = 5/5$.

Random walk examples.

Suppose you start with 2 chips at time 0 and that your tournament is like a simple random walk, but if you hit 0 you are done.

a. P(you have not hit zero by time 4)? b. P(you have not hit zero by time 46)?

a. We can just count here. There are $2^4 = 16$ paths of length 4, each equally likely.

+ + + +, + + + -, + + - +, etc.

How many hit zero? - - + +, - - + -, - - - +, - - - -, + - - -, - + - -.

The other 10/16 we avoid zero. So the answer is 10/16 = 62.5%.

b. We saw last time how starting with 1 chip at time 0, $P(Y_1 > 0, Y_2 > 0, \ldots, Y_{47} > 0)$
$= \text{Choose}(48,24)(½)^{48} = 11.46\%$.

For this to happen, you have to win your first hand and go up to 2 chips. Therefore

$P(Y_1 > 0, Y_2 > 0, \ldots, Y_{47} > 0) = P(Y_1 = 2, Y_2 > 0, \ldots, Y_{47} > 0)$

$= ½$ P(starting with 2 chips, $Y_1 > 0, Y_2 > 0, \ldots, Y_{46} > 0$).

So starting with 2 chips, $P(Y_1 > 0, Y_2 > 0, \ldots, Y_{46} > 0) = 2\ (11.46\%) = 22.92\%$.

Tournaments.